

# NFDI-Datenspeicher Goethe Universität

V. Lindenstruth, A. Redelbach

28.5.2025

# Goethe NHR

Moved to Green IT cube at GSI/FAIR

Number	Type	Vendor	CPU	GPU	Cores per CPU	Cores per Node	Threads per Node	RAM [GB]
412	dual-socket	Intel	Xeon Skylake Gold 6148	n/a	20	40	80	192
72	dual-socket	Intel	Xeon Skylake Gold 6148	n/a	20	40	80	772
139	dual-socket	Intel	Xeon Broadwell E5-2640 v4	n/a	10	20	40	128
112	dual-socket	AMD	EPYC 7452	8x MI210	32	64	128	512



HPC system Goethe-NHR:

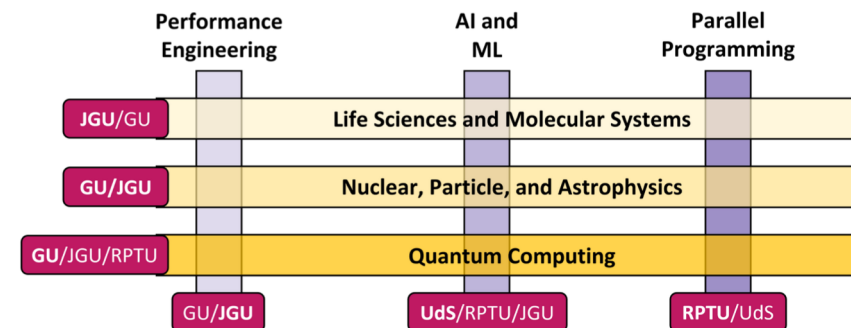
Ranked 34 in Green 500 list (46,5 GigaFlops/Watt), **Rank 9 in June 2023**

Ranked 131 in Top 500 list (9 PFlop/s)

NHR South-West:

<https://nhrsw.de/about/>

## Core and Research Competencies



# Goethe University data repository

The archiving and publication platform for scientific research data at Goethe University Frankfurt.

The Goethe University Data Repository (GUDe) provides a platform for its members to electronically archive, share, and publish their research data. GUDe is jointly operated by the University Library and the University Data Center of the Goethe University.

**Metadata** of all public content is freely available and indexed by search engines as well as scientific web services. GUDe follows the FAIR principles for long-term accessibility (**minimum 10 years**), allows for reliable **citation via DOIs** as well as cooperative access to non-public data and operates on DSpace-CRIS v7.

<https://docs.gude.uni-frankfurt.de/step-by-step-guide/>

<https://docs.gude.uni-frankfurt.de/documentation/>



# New mass storage at Goethe University

High-performance storage at large scale, exceptional storage performance for different workloads

Hybrid architecture of **VDURA**, meeting the needs of HPC/AI workflows

Nodes are directly connected to the InfiniBand network

**PanFS parallel file system** for data/metadata

de-clustered erasure coding: data is broken into blocks and stored across multiple physical locations

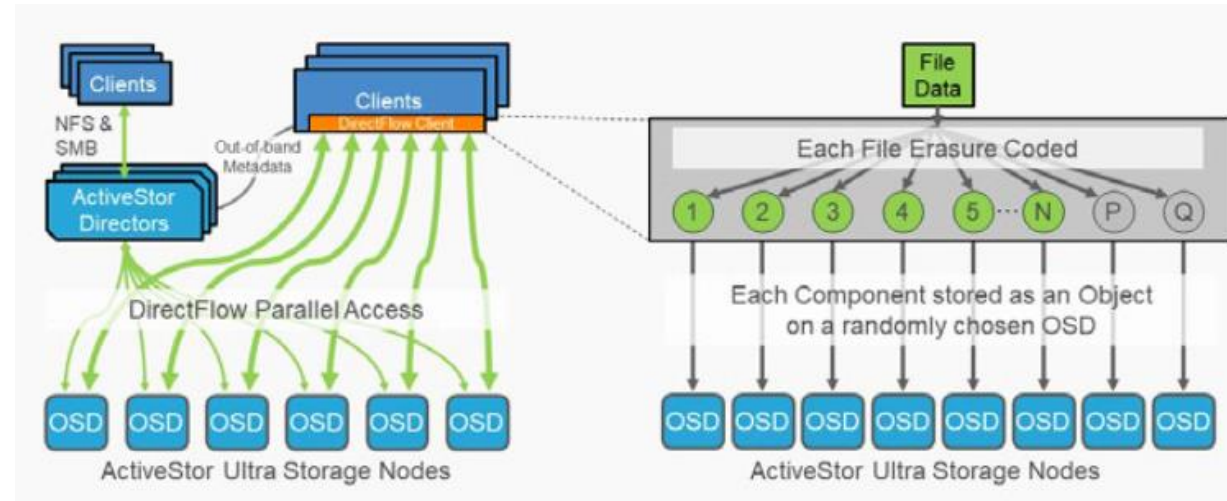
new **multi-level erasure coding**: each storage server will perform erasure coding within its own storage elements

Performant – guaranteed performance 90 GB/s reading and writing

Economical – lowest TCO (total cost of operation) of any parallel file system

# New mass storage at Goethe University

*VDURA parallel architecture  
with Object Storage Devices (OSDs)*



Main components:

**3 director nodes**, each node features 12 3.8 TB NVMe SSD

director nodes provide high-available metadata and gateway services for protocols like NFS, SMB and S3

**7 storage nodes**, aiming at high performance for mixed file size, HPC, and AI/ML workloads

Each storage node consists of a 1 unit head node connected to a single 4 unit 108-bay JBOD, racks fully populated with HDDs

**DirectFlow client driver**: SW module that runs on Linux compute servers ("Clients") and interacts with director/storage nodes to read/write the files stored by PanFS.

- usable capacity is 15.8 PB
- POSIX- compliant file system
- Architecture for linear scale-out

# Requirements of PUNCH4NFDI/Storage4PUNCH

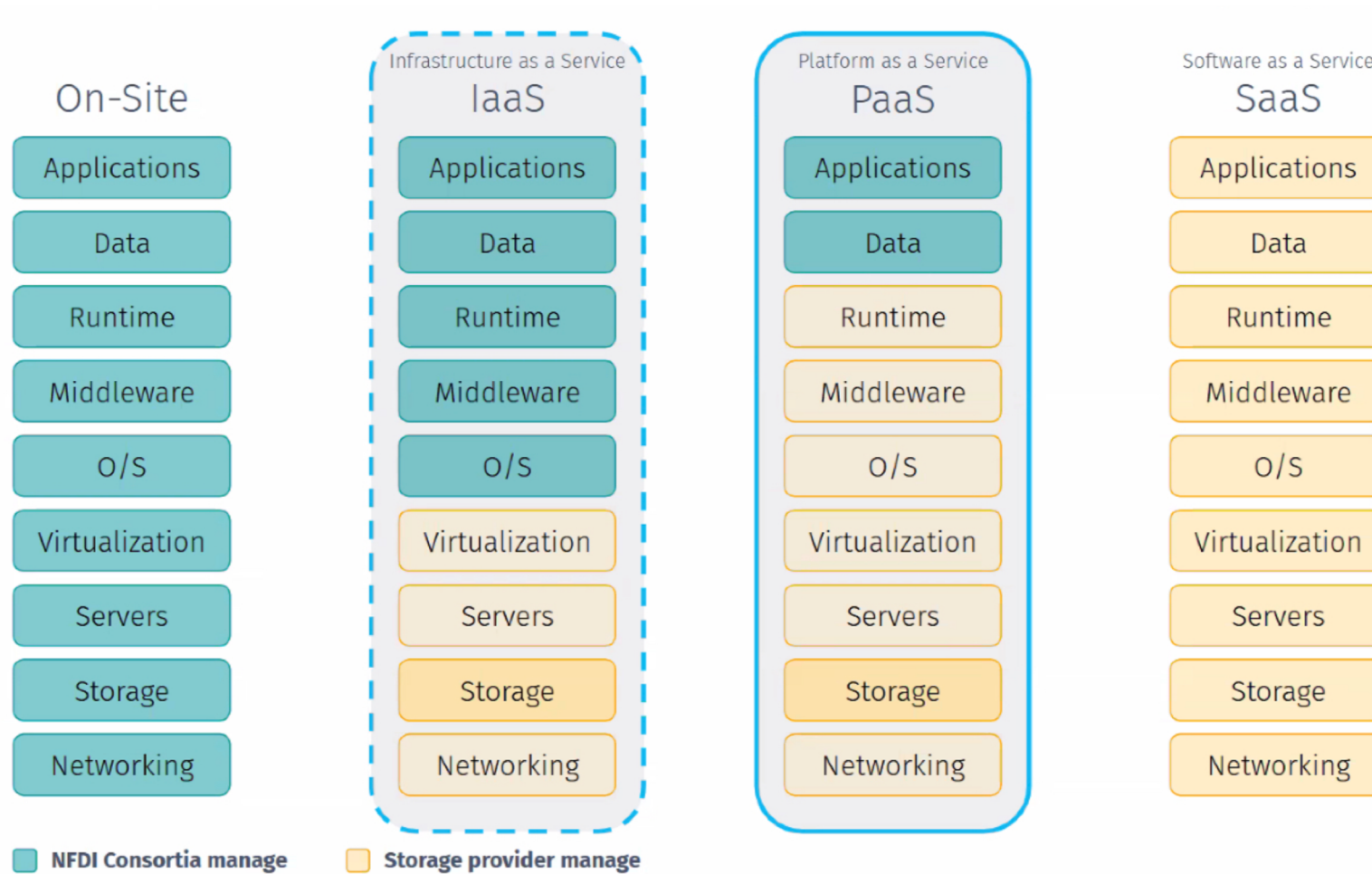
Estimate for storage  
for the next 5 years

		HS 1	HS 2	HS 3	...	Bereits realisiert	offen
		beantragt	beantragt	beantragt			
NFDI-Konsortium 1	Artikulierte Bedarf und Planungen	...PB (...%)				...PB (...%)	...PB (...%)
NFDI-Konsortium 2	dito		...PB (...%)	...PB (...%)		...	...
NFDI-Konsortium 3	dito			...PB (...%)		...	...
...	...	...	...	...		...	...
		$\Sigma$ : ...PB	$\Sigma$ : ...PB	$\Sigma$ : ...PB			

Draft of WG Overall Architecture:

PaaS	HPC CPU	Für Analyse	Angebot besteht durch NHR / GCS Anbindung an Storage erforderlich
PaaS	HPC GPU	Für ML/KI	
PaaS	(Semi-)Structured Storage (SQL, NoSQL, Graph, ...)	Zugriffsmuster: Random Access, Heiß, Warm, Kalt, Archiv (WORM/N) Daten: Datenmenge, Dateianzahl, Datentypen (wenn bekannt) Persistenz: scratch, 10y, LTA Speicherbedarf auf der Zeitachse für 5 Jahre	Storage-Providers
PaaS	Volume Storage (POSIX, Objekt, ...)		
PaaS	Kubernetes Cluster	Für Analyse / aufbauende Services (CPU/GPU?)	Storage-Providers (notwendig für Speicher- und Dienstangebot aber nicht im Fokus des DFG Calls)
IaaS	Virtual Machines		

# Cloud service model as basis (PAAS/IAAS)



Platform integrating storage



# Coordination with NFDI / other consortia

## **Timeline:**

Ende Mai 2025: Veröffentlichung der Ausschreibung

Juni / Juli 2025: Informationsveranstaltung der DFG für Rechenzentren und NFDI-Konsortien

August 2025: Eingang Absichtserklärungen

Oktober 2025: Eingang Anträge

Februar 2026: Förderentscheidungen

## **Governance**

## **Connections to other consortia**